

AD-A238 689



ARO 27868.7-MA

(2)

A CONSTRUCTIVE DEFINITION OF DIRICHLET PRIORS

Jayaram Sethuraman†

Department of Statistics
Florida State University
Tallahassee, Florida 32306-3033

FSU Technical Report No. M-843
USARO Technical Report No. D-118

May 1991

SEARCHED	INDEXED
SERIALIZED	FILED
MAY 1991	
FBI - TAMPA	
A-1	

† Research supported by U. S. Army Research Office Grant DAAL03-90-G-0103.
AMS (1980) subject classification nos. Primary 60G57, 62A15; secondary 62G99, 62E99.

Key words and phrases: Bayesian Nonparametrics, Random Probability Measures, Dirichlet Measures.

91-05395



91 7 1 136

A CONSTRUCTIVE DEFINITION OF DIRICHLET PRIORS

By Jayaram Sethuraman

Abstract

The "parameter" in a Bayesian nonparametric problem is the unknown distribution P of the observation X . A Bayesian uses a prior distribution for P , and after observing X , solves the statistical inference problem by using the posterior distribution of P , which is the conditional distribution of P given X . For Bayesian nonparametrics to be successful one needs a large class of priors for which posterior distributions can be easily calculated.

Unless X takes values in a finite space, the unknown distribution P varies in an infinite dimensional space. Thus one has to talk about measures in a complicated space like the space of all probability measures on a large space. This has always required a more careful attention to the attendant measure theoretic problems.

A class of priors known as Dirichlet measures have been used for the distribution of a random variable X when it takes values in \mathcal{R}_k , see Freedman (1963), Fabius (1964) and Ferguson (1973). This family forms a conjugate family and possesses many pleasant properties.

In this paper we give a simple and new constructive definition of Dirichlet measures and remove the restriction that the basic space should be \mathcal{R}_k . We give complete self contained proofs of the three basic results for Dirichlet measures:

1. The Dirichlet measure is a probability measure of on the space of all probability measures,
2. it gives probability one to the subset of discrete probability measures, and
3. the posterior distribution is also a Dirichlet measure.

1. Introduction.

Bayesian nonparametrics came into vogue in the seventies. Let X be a random variable taking values in a measurable space $(\mathcal{X}, \mathcal{B})$ and let its unknown probability measure be P . The “parameter” in a Bayesian nonparametrics problem is the unknown probability distribution P . If \mathcal{X} is not a finite set, this parameter takes values in an infinite dimensional space, and hence the definition of a prior distribution for P has always required a more careful description of the attendant measure theoretic problems. A practitioner of Bayesian nonparametrics puts a prior distribution for P and gives his answer to the inference problem as the posterior distribution of P given X . How do we define such a prior distribution and calculate the posterior distribution? Let \mathcal{P} be the space of probability measures on $(\mathcal{X}, \mathcal{B})$ and note that P varies in \mathcal{P} . A natural σ -field in \mathcal{P} is \mathcal{C} , the smallest σ -field generated by sets of the form $\{P : P(B) < r\}$ where B varies in \mathcal{B} and r varies in $[0, 1]$. A nonparametric prior for a probability measure P is then a probability measure ν on $\{\mathcal{P}, \mathcal{C}\}$. Let (P, X) be a pair of random variables taking values in $\{\mathcal{P} \times \mathcal{X}, \mathcal{C} \times \mathcal{B}\}$ such that P has distribution ν and such that X given P has distribution P . The posterior distribution ν^X is defined to be the distribution of P given X .

Bayesian nonparametrics becomes tractable only if there are examples of priors ν for which ν^X are easy to calculate. A collection of prior distributions ν_α indexed by a parameter α is said to form a conjugate family of priors if the posterior distribution ν_α^X is of the form $\nu_{f(\alpha, X)}$ for some function $f(\alpha, X)$ of α and X . The class of Dirichlet measures form a conjugate family that makes it useful in Bayesian nonparametrics.

Before giving an intuitive definition of a Dirichlet measure we will repeat the well known definition of Dirichlet measures on finite dimensional spaces. Let $(\gamma_1, \gamma_2, \dots, \gamma_k)$ be a vector such that $\gamma_j \geq 0, j = 1, 2, \dots, k$ and such that $\sum \gamma_j > 0$. Let $z_{\gamma_j}, j = 1, 2, \dots, k$ be independent Gamma random variables with scale parameter 1 and shape parameters $\gamma_j, j = 1, 2, \dots, k$, respectively. Let $z = \sum z_{\gamma_j}$ and $y_j = (z_{\gamma_j}/z), j = 1, 2, \dots, k$. The joint distribution of the random variable (y_1, y_2, \dots, y_k) taking values in $\mathcal{P}_k = \{(p_1, p_2, \dots, p_k) : p_1 \geq 0, p_2 \geq 0, \dots, p_k \geq 0, \sum p_j = 1\}$, the unit simplex of \mathcal{R}_k , is defined to be k -dimensional Dirichlet measure, $\mathcal{D}_{(\gamma_1, \gamma_2, \dots, \gamma_k)}$. Let e_j denote the k -dimensional vector consisting of 0's, except for the j th co-ordinate, which is equal to 1. Notice that the Dirichlet measure \mathcal{D}_{e_j} puts all its probability mass at the point e_j . Further more, it is interesting to note that $\mathcal{D}_{2e_j} = \mathcal{D}_{e_j}$. This fact will be used later in the proof of Theorem 4.3.

The intuitive definition of a Dirichlet measure in the general case is easy to give. Let α be a non-zero element of \mathcal{M} , i.e. let α be a non-zero finite measure on $(\mathcal{X}, \mathcal{B})$. A probability distribution ν on $(\mathcal{P}, \mathcal{C})$ is said to be a Dirichlet measure with parameter α if for every measurable partition $\{B_1, B_2, \dots, B_k\}$ of \mathcal{X} , the distribution of $(P(B_1), P(B_2), \dots, P(B_k))$ under ν is the finite dimensional Dirichlet distribution $\mathcal{D}_{(\alpha(B_1), \alpha(B_2), \dots, \alpha(B_k))}$. When such a probability measure ν on $(\mathcal{P}, \mathcal{C})$ can be demonstrated to exist, it will be denoted by \mathcal{D}_α .

There are three main properties of Dirichlet measures that make them useful in Bayesian nonparametrics. Apart from their marginals having finite dimensional Dirichlet distributions, they possess the following three properties:

- P1** \mathcal{D}_α is a probability measure on $(\mathcal{P}, \mathcal{C})$,
- P2** \mathcal{D}_α gives probability one to the subset of all discrete probability measures on $(\mathcal{X}, \mathcal{B})$, and
- P3** the posterior distribution \mathcal{D}_α^X is the Dirichlet measure $\mathcal{D}_{\alpha + \delta_X}$ where δ_X is the probability measure degenerate at X . This paper gives a constructive definition of a Dirichlet measure and shows that these three properties hold.

Ferguson (1973) argued that the distributions of $(P(B_1), P(B_2), \dots, P(B_k))$ gave rise to a consistent family of measures over the class of all partitions (B_1, B_2, \dots, B_k) . By the Kolmogorov consistency theorem this gives rise to a unique probability measure on $[0, 1]^{\mathcal{B}}$ with its associated Kolmogorov σ -field. Further more, for any given sequence of disjoint measurable sets B_1, B_2, \dots , the probability is one that

$$P(\cup B_j) = \sum P(B_j), \quad (1.1)$$

where $P(\cdot)$ is the canonical representation of a point in $[0, 1]^{\mathcal{B}}$. This set of probability one may depend on the sequence B_1, B_2, \dots . Such a P is a member of \mathcal{P} if and only if (1.1) were true for all disjoint sequences B_1, B_2, \dots . The collection of such disjoint sequences is uncountable. This presents a problem in making this definition rigorous and establishing property **P1**. For the special case where \mathcal{X} is the real line, or more generally a separable complete metric space, one can use a result of Harris (1968, Lemma 6.1). This result states that a verification of (1.1) for a select countable number of cases of disjoint sequences of sets is sufficient to ensure that (1.1) holds for all disjoint countable sets and that the set function P is a probability measure. An appeal to this result is one way to show that there is a probability measure on $(\mathcal{P}, \mathcal{C})$ with the required properties and this defines the Dirichlet measure \mathcal{D}_α .

In a later section, Ferguson ((1973), Section 4) gives an alternative constructive

definition of the Dirichlet measure which shows that it gives probability one to the subset of discrete probability measures. However, it takes some effort to see that that the two definitions are equivalent.

Ferguson (1973) also establishes the posterior distribution property **P3** by using a very peculiar definition (see his Definition 2) for the joint distribution of (P, X) .

Blackwell and McQueen (1973) appeal to the famous theorem of de Finetti to show that there is a one-to-one correspondence between sequences of exchangeable random variables and probability measures on $(\mathcal{P}, \mathcal{C})$. A particular case of exchangeable random variables, namely the generalized Pòlya urn scheme, corresponds to the Dirichlet measure. In this paper and in Blackwell (1973), they establish the three properties **P1**, **P2** and **P3**. Their proof is elegant but quite indirect and also requires the space \mathcal{X} to be a separable complete metric space.

Freedman (1963) and Fabius (1964) contain early work on tail-free priors, which include Dirichlet priors, for the case when \mathcal{X} is the set of integers or $[0, 1]$.

Let \mathcal{E} be the usual Borel σ -field restricted to $[0, 1]$. In Section 2, we define a function P based on a sequence of i.i.d. random variables $(\theta_n, Y_n), n = 1, 2, \dots$ taking values in $([0, 1] \times \mathcal{X}, \mathcal{E} \times \mathcal{B})$. See (2.1). By its very definition, P is a random measure taking values in $(\mathcal{P}, \mathcal{C})$ and giving probability one to the subset of discrete probability measures on $(\mathcal{X}, \mathcal{B})$. This establishes properties **P1** and **P2**. We give a direct proof, in Theorem 3.4 of Section 3, that the finite dimensional marginal distributions of P are Dirichlet distributions. This establishes that the distribution of P is a Dirichlet measure. In Theorem 4.3 of Section 4 we prove property **P3** thus establishing that the posterior distribution is also a Dirichlet measure. The definition and proofs are all given in some detail to make this paper self contained.

This constructive definition of a Dirichlet measure was announced in a paper on convergence of Dirichlet measures, Sethuraman and Tiwari (1982). This definition has since been used by several authors to greatly simplify previous calculations and to obtain new calculations involving Dirichlet measures. For instance see Ferguson (1983), Ferguson, Phadia and Tiwari (1991), Kumar and Tiwari (1989).

2. Constructive definition of the Dirichlet measure

Let α be a non-zero finite measure on $(\mathcal{X}, \mathcal{B})$. Let $\beta(B) = \alpha(B)/\alpha(\mathcal{X})$ be the

normalized probability measure arising from α . Let $B(\gamma, \delta)$ stand for the Beta distribution on $[0, 1]$ with parameters γ and δ . This Beta distribution is the marginal distribution of the first co-ordinate of the Dirichlet measure $\mathcal{D}(\gamma, \delta)$ on the two-dimensional simplex \mathcal{P}_2 defined earlier. Let $\mathcal{N} = \{1, 2, \dots\}$ be the set of positive integers and let \mathcal{F} be the σ -field of all subsets of \mathcal{N} . Let $\{\Omega, \mathcal{S}, Q\}$ be a probability space supporting a collection of random variables $(\boldsymbol{\theta}, \mathbf{Y}, I) = ((\theta_j, Y_j), j = 1, 2, \dots, I)$ taking values in $(([0, 1] \times \mathcal{X})^\infty \times \mathcal{N}, (\mathcal{E} \times \mathcal{B})^\infty \times \mathcal{F})$, with a joint distribution defined as follows. The random variables $(\theta_1, \theta_2, \dots)$ are i.i.d. with a common Beta distribution $B(1, \alpha(\mathcal{X}))$. The random variables (Y_1, Y_2, \dots) are independent of the $(\theta_1, \theta_2, \dots)$ and i.i.d. among themselves with common distribution β . Let $p_1 = \theta_1$ and for $p_n = \theta_n \prod_{1 \leq m < n-1} (1 - \theta_m)$ for $n = 2, 3, \dots$. Notice that $\sum_{1 \leq m \leq n} p_m = 1 - \prod_{1 \leq m \leq n} (1 - \theta_m) \rightarrow 1$ with Q -probability one. Let $Q(I = n | (\boldsymbol{\theta}, \mathbf{Y})) = p_n, n = 1, 2, \dots$. The existence of a probability space (Ω, \mathcal{S}, Q) and such a sequence of random variables $(\boldsymbol{\theta}, \mathbf{Y}, I)$ follows from the usual construction of a product measure, and does not require any restrictions on $(\mathcal{X}, \mathcal{B})$, such as its being a separable complete metric space.

Define

$$P(\boldsymbol{\theta}, \mathbf{Y}; B) = P(B) = \sum_{n=1}^{\infty} p_n \delta_{Y_n}(B) \quad (2.1)$$

where $\delta_x(\cdot)$ stands for the probability measure degenerate at x .

This is the new constructive definition of a Dirichlet measure. As convenience dictates, we drop all or part of the arguments $(\boldsymbol{\theta}, \mathbf{Y}), B$ and denote the random measure in (2.1) by P , for simplicity of notation. Since P is clearly a measurable map from (Ω, \mathcal{S}) into $(\mathcal{P}, \mathcal{C})$ and takes values in the subset of discrete probability measures, properties **P1** and **P2** are self evident.

Notice that the random variable I introduced above has not been used in the definition of P . It will be used later, in Section 4, to prove the posterior distribution property **P3**.

A more direct way to describe the constructive definition in (2.1) is as follows. Let Y_1, Y_2, \dots be i.i.d. with common distribution β . Let $\{p_1, p_2, \dots\}$ be the probabilities from a discrete distribution on the integers with discrete failure rate $\{\theta_1, \theta_2, \dots\}$ which are i.i.d. with a Beta distribution $B(1, \alpha(\mathcal{X}))$. Let P be the random probability measure that puts weights p_n at the degenerate measures δ_{Y_n} , $n = 1, 2, \dots$. This is the random probability measure P described in (2.1). The alter-

native definition given in Ferguson ((1973), Section 4) uses a different set of random weights which are arranged in decreasing order. The use of unordered weights in this paper simplifies all our calculations. It is interesting to note that the weights used by Ferguson (1973) are equivalent to our weights rearranged in decreasing order. However, it is not clear that there is an easy way to unordered the weights of Ferguson (1973) to obtain weights with the simple structure of (2.1).

3. The distribution of the random measure P is \mathcal{D}_α .

We will digress a little before establishing that the distribution of P is the Dirichlet measure \mathcal{D}_α .

Let $\theta_n^* = \theta_{n+1}$, $Y_n^* = Y_{n+1}$, $n = 1, 2, \dots$ and let $J = I - 1$. Define $(\theta^*, Y^*, J) = ((\theta_1^*, \theta_2^*, \dots), (Y_1^*, Y_2^*, \dots), J)$.

Notice that

$$P(\theta, Y; B) = \theta_1 \delta_{Y_1}(B) + (1 - \theta_1) P(\theta^*, Y^*; B). \quad (3.1)$$

Notice that (θ^*, Y^*) has the same distribution as (θ, Y) and is independent of (θ_1, Y_1) . Thus we can re-write (3.1) as the following distributional equation for P :

$$P \stackrel{\text{st}}{=} \theta_1 \delta_{Y_1} + (1 - \theta_1) P, \quad (3.2)$$

where on the right hand side P is independent of (θ_1, Y_1) .

Theorem 3.4 below uses the distributional equation (3.2) to show that the distribution of P is the Dirichlet measure \mathcal{D}_α . The proof of this theorem uses well known facts about finite dimensional Dirichlet measures and a result on the uniqueness of solutions to distributional equations, which are given below as Lemmas 3.1, 3.2 and 3.3.

Lemma 3.1 *Let $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_k)$ and $\delta = (\delta_1, \delta_2, \dots, \delta_k)$ be k -dimensional vectors. Let U, V be independent k -dimensional random vectors with Dirichlet distributions \mathcal{D}_γ and \mathcal{D}_δ , respectively. Let W be independent of (U, V) and have a Beta distribution $B(\gamma, \delta)$, where $\gamma = \sum \gamma_j$ and $\delta = \sum \delta_j$. Then the distribution of $WU + (1 - W)V$ is the Dirichlet distribution $\mathcal{D}_{\gamma + \delta}$.*

Lemma 3.2 Let $\gamma = (\gamma_1, \dots, \gamma_k)$, $\gamma = \sum \gamma_j$ and let $\beta_j = \gamma_j/\gamma, j = 1, 2, \dots, k$. Then

$$\sum \beta_j \mathcal{D}_{\gamma + e_j} = \mathcal{D}_\gamma.$$

The proofs of these two lemmas are found in many standard text books, for instance in Wilks ((1962), Section 7).

Lemma 3.3 stated and proved below shows that certain distributional equations have unique solutions. Such results appear in several areas of statistics, notably in renewal theory. For a recent work which gives more general results see Goldie (1991). The following lemma is sufficient for our purposes. Its proof, which is not new, is given here to make this paper self contained.

Lemma 3.3 Let W, U, V be random variables where W is a real valued and U, V take values in a linear space. Suppose that V is independent of (W, U) and satisfies the distributional equation

$$V \stackrel{st}{=} U + WV. \quad (3.3)$$

Suppose that $P(W = 1) \neq 1$. Then there is only one distribution for V that satisfies (3.3).

Proof: Let V and V' be two random variables whose distributions are not equal but satisfy equation (3.3). Let (W_n, U_n) be independent copies of (W, U) which are independent of V, V' . Let $V_1 = V, V'_1 = V'$ and define, recursively,

$$V_{n+1} = U_n + W_n V_n \text{ and } V'_{n+1} = U_n + W_n V'_n$$

for $n = 1, 2, \dots$. From the distributional equation (3.3), the V_n 's have the same distribution as V and the V'_n 's have the same distribution as V' . However,

$$|V_{n+1} - V'_{n+1}| = |W_n| |V_n - V'_n| = \prod_{1 \leq m \leq n} |W_m| |V_1 - V'_1| \rightarrow 0$$

with probability 1, since the W_n 's are i.i.d. and $P(W = 1) < 1$. This contradicts the supposition that the distributions of V and V' are unequal and proves that the distribution of V satisfying (3.3) is unique. \diamond

Theorem 3.4 Let $\{B_1, B_2, \dots, B_k\}$ be a measurable partition of \mathcal{X} and let $\mathbf{P} = (P(B_1), P(B_2), \dots, P(B_k))$. Then the distribution of \mathbf{P} is the k -dimensional Dirichlet measure $\mathcal{D}_{(\alpha(B_1), \alpha(B_2), \dots, \alpha(B_k))}$.

Proof: Let $\mathbf{D} = (\delta_{Y_1}(B_1), \delta_{Y_1}(B_2), \dots, \delta_{Y_1}(B_k))$. Notice that $P(\mathbf{D} = \mathbf{e}_j) = P(Y_1 \in B_j) = \beta(B_j), j = 1, 2, \dots, k$. From (3.2) we see that \mathbf{P} satisfies the distributional equation

$$\mathbf{P} \stackrel{st}{=} \theta_1 \mathbf{D} + (1 - \theta_1) \mathbf{P}, \quad (3.4)$$

where, on the right, θ_1 has a Beta distribution $B(1, \alpha(\mathcal{X}))$, \mathbf{D} is independent of θ_1 and takes the value \mathbf{e}_j with probability $\beta(B_j), j = 1, 2, \dots, k$, and the k -dimensional random vector \mathbf{P} is independent of (θ_1, \mathbf{D}) .

We will first verify that the k -dimensional Dirichlet measure for \mathbf{P} satisfies the distributional equation (3.4) and then show that this solution is the unique solution.

Let the distribution of \mathbf{P} on the right of (3.4) be the k -dimensional Dirichlet measure $\mathcal{D}_{(\alpha(B_1), \alpha(B_2), \dots, \alpha(B_k))}$. The k -dimensional Dirichlet measure $\mathcal{D}_{\mathbf{e}_j}$ gives probability 1 to \mathbf{e}_j . Given that $\mathbf{D} = \mathbf{e}_j$, the distribution of $\theta_1 \mathbf{D} + (1 - \theta_1) \mathbf{P}$ is the distribution of $\theta_1 \mathcal{D}_{\mathbf{e}_j} + (1 - \theta_1) \mathcal{D}_{(\alpha(B_1), \alpha(B_2), \dots, \alpha(B_k))}$ and this, by Lemma 3.1, is $\mathcal{D}_{(\alpha(B_1), \alpha(B_2), \dots, \alpha(B_k)) + \mathbf{e}_j}$. Summing over the distribution of \mathbf{D} is equivalent to taking a mixture of these Dirichlet measures with weights $\beta(B_j) = \alpha(B_j)/\alpha(\mathcal{X})$, which by Lemma 3.2, is equal to $\mathcal{D}_{(\alpha(B_1), \alpha(B_2), \dots, \alpha(B_k))}$. This verifies that the k -dimensional Dirichlet measure satisfies the distributional equation (3.4). Lemma 3.3 shows that this solution is unique. This completes the proof of Theorem 3.4. \diamond

4. The posterior distribution of P is $\mathcal{D}_{\alpha + \delta_X}$.

Let $X = Y_I$. Then X is a random variable from (Ω, \mathcal{S}) into \mathcal{X} defined explicitly as a function of $(\boldsymbol{\theta}, \mathbf{Y}, I)$. The next lemma shows that the distribution of X given P is P and hence the joint distribution of (P, X) is that of the parameter and observation in a Bayesian nonparametric problem.

Lemma 4.1 *The distribution of X given P is P .*

Proof: Let $B \in \mathcal{B}$. By direct calculation, we get

$$\begin{aligned} Q(X \in B | (\boldsymbol{\theta}, \mathbf{Y})) &= \sum_n Q(X \in B, I = n | (\boldsymbol{\theta}, \mathbf{Y})) Q(I = n | (\boldsymbol{\theta}, \mathbf{Y})) \\ &= \sum_n Q(Y_n \in B | (\boldsymbol{\theta}, \mathbf{Y})) p_n \\ &= \sum_n p_n \delta_{Y_n}(B) = P(B). \end{aligned}$$

Since this conditional probability is a function of P , it immediately follows that $Q(\cdot|P)$ exists as a regular conditional probability and $Q(X \in B|P) = P(B)$ with Q -probability 1. \diamond

We now come to the posterior distribution of P , i.e. the distribution of P given X . We do this by separately obtaining the conditional distribution of (θ, Y) given $I = 1$ and given $I > 1$. When f and g are functions of (θ, Y, I) , we will use the notations $\mathcal{L}(f)$ and $\mathcal{L}(f|g)$ to denote the distribution of f and the conditional distribution of f given g , under Q , respectively.

Lemma 4.2 *The following are the conditional distributions of (θ, Y, I) given $I = 1$ and given $I > 1$:*

$$\mathcal{L}((\theta_1, Y_1), (\theta^*, Y^*)|I = 1) = B(2, \alpha(\mathcal{X})) \times \mathcal{L}(\theta, Y) \quad (4.1)$$

and

$$\mathcal{L}((\theta_1, Y_1), (\theta^*, Y^*), J|I > 1) = B(1, \alpha(\mathcal{X}) + 1) \times \mathcal{L}(\theta, Y, I). \quad (4.2)$$

Proof: Notice that $Q(I = 1|(\theta, Y)) = \theta_1$. Thus, if $A_i \in \mathcal{E}, B_i \in \mathcal{B}, i = 1, 2, \dots, n$, we have the relation

$$\begin{aligned} & Q\{\theta_i \in A_i, Y_i \in B_i, i = 1, 2, \dots, n, I = 1\} \\ & \propto \int I(x_i \in A_i, y_i \in B_i, i = 1, 2, \dots, n) x_1 \prod_{1 \leq i \leq n} [(1 - x_i)^{\alpha(\mathcal{X})-1} dx_i \beta(dy_i)]. \end{aligned}$$

This implies, conditional on $I = 1$, θ_1 has distribution $B(2, \alpha(\mathcal{X}))$, the distributions of $\theta_i, i = 2, 3, \dots, n$ and $Y_i, i = 1, 2, \dots, n$ are all unchanged, and all these are independent. This gives all the finite dimensional conditional distributions and proves (4.1). The proof of (4.2) follows along the same lines since $Q(I > 1|(\theta, Y)) = 1 - \theta_1$. \diamond

Theorem 4.3 *The posterior distribution of P given X is the Dirichlet measure $\mathcal{D}_{\alpha+\delta_X}$.*

Proof: Let $P^* = P(\theta^*, Y^*)$. We can rewrite (3.1) as

$$P = \theta_1 \delta_{Y_1} + (1 - \theta_1) P^*. \quad (4.3)$$

When $I = 1$, we use (4.1) and obtain

$$\begin{aligned}\mathcal{L}(P|X, I = 1) &= \mathcal{L}(\theta_1 \delta_{Y_1} + (1 - \theta_1) P^* | X, I = 1) \\ &\stackrel{st}{=} \theta'_1 \delta_X + (1 - \theta'_1) P^{**}\end{aligned}\quad (4.4)$$

where θ' has distribution $B(2, \alpha(\mathcal{X}))$, and P^{**} is a random probability measure, independent of θ' , whose distribution is the Dirichlet measure \mathcal{D}_α . The random probability measure putting all its mass on the degenerate measure δ_X is the Dirichlet measure \mathcal{D}_{δ_X} which is also equal to $\mathcal{D}_{2\delta_X}$. Since θ' has a Beta distribution $B(2, \alpha(\mathcal{X}))$, this latter choice allows us to use Lemma 3.1 to obtain

$$\mathcal{L}(P|X, I = 1) \stackrel{st}{=} \mathcal{D}_{\alpha+2\delta_X}. \quad (4.5)$$

When $I > 1$, we use (4.2) and first obtain

$$\mathcal{L}(\theta^*, Y^*, X | I > 1) = \mathcal{L}(\theta, Y, X) \quad (4.6)$$

since $X = Y_I = Y_J^*$ on $I > 1$. Thus

$$\begin{aligned}\mathcal{L}(P|X, I > 1) &= \mathcal{L}(\theta_1 \delta_{Y_1} + (1 - \theta_1) P^* | X, I > 1) \\ &\stackrel{st}{=} \theta''_1 \delta_{Y_1} + (1 - \theta''_1) P^{***}\end{aligned}\quad (4.7)$$

where Y_1 has distribution β , θ''_1 is independent of Y_1 and has distribution $B(1, \alpha(\mathcal{X}) + 1)$, and P^{***} is a random probability measure, independent of (Y_1, θ''_1) , whose distribution is $\mathcal{L}(P|X)$, in view of (4.6). We can combine (4.4) and (4.7) to obtain a distributional equation for $\mathcal{L}(P|X)$ as follows.

$$\mathcal{L}(P|X) \stackrel{st}{=} A(\theta'_1 \delta_X + (1 - \theta'_1) P^{**}) + (1 - A)(\theta''_1 \delta_{Y_1} + (1 - \theta''_1) P^{***}), \quad (4.8)$$

where all the random variables on the right are independent and have the distributions previously specified, and the random variable A takes values 1 and 0 with probabilities $\frac{1}{(\alpha(\mathcal{X})+1)}$ and $\frac{\alpha(\mathcal{X})}{(\alpha(\mathcal{X})+1)}$, respectively. Notice that the distribution of P^{***} is $\mathcal{L}(P|X)$ which makes (4.8) a distributional equation.

From Lemma 3.3 we conclude that if there is a solution to (4.8), it will be a unique solution. We will now verify that $\mathcal{L}(P|X) = \mathcal{D}_{\alpha+\delta_X}$ verifies the distributional equation (4.8). Relation (4.5) can be rewritten as

$$\theta'_1 \delta_X + (1 - \theta'_1) P^{**} \stackrel{st}{=} \mathcal{D}_{\alpha+2\delta_X}. \quad (4.9)$$

By conditioning on Y_1 and using Lemma 3.1, and then taking expectations with respect to Y_1 , we find that

$$\theta_1'' \delta_{Y_1} + (1 - \theta_1'') P^{***} \stackrel{st}{=} E(\mathcal{D}_{\alpha + \delta_X + \delta_{Y_1}}), \quad (4.10)$$

where Y_1 has distribution β . Let Z be a random variable in $(\mathcal{X}, \mathcal{B})$ with distribution $\frac{1}{(\alpha(\mathcal{X})+1)} \delta_X + \frac{\alpha(\mathcal{X})}{(\alpha(\mathcal{X})+1)} \beta = \frac{\alpha + \delta_X}{(\alpha(\mathcal{X})+1)}$. Combining (4.9) and (4.10), and using Lemma 3.2 on mixtures of Dirichlet measures, we conclude the distribution of the random measure in the right hand side of (4.8) is equal to

$$\begin{aligned} \frac{1}{(\alpha(\mathcal{X})+1)} \mathcal{D}_{\alpha+2\delta_X} + \frac{\alpha(\mathcal{X})}{(\alpha(\mathcal{X})+1)} E(\mathcal{D}_{\alpha+\delta_X+\delta_{Y_1}}) &\stackrel{st}{=} E(\mathcal{D}_{\alpha+\delta_X+\delta_Z}) \\ &\stackrel{st}{=} \mathcal{D}_{\alpha+\delta_X}. \end{aligned}$$

This proves that $\mathcal{D}_{\alpha+\delta_X}$ is the posterior distribution of P given X . ◇

5. Acknowledgement

Professor David Blackwell was visiting Florida State University in Spring 1979 when I was giving a seminar on Bayesian nonparametrics. It was then that I discovered this new construction of a Dirichlet measure. I wish to thank Professor Blackwell for his inspiration and encouragement.

6. References

- Blackwell, D. (1973) Discreteness of Ferguson priors, *Ann. Statist.* **1** 356–358.
- Blackwell, D. and MacQueen, J. B. (1973) Ferguson distributions via Pólya schemes, *Ann. Statist.* **1** 353–355.
- Fabius, J. (1964) Asymptotic behavior of Bayes' estimates, *Ann. Math. Statist.* **35** 846–856.
- Ferguson, T. S. (1973) A Bayesian analysis of some nonparametric problems, *Ann. Statist.* **1** 209–230.
- Ferguson, T. S. (1983) A Bayesian density estimation by mixtures of normal distributions, *Recent Advances in Statistics*, Academic Press.
- Ferguson, T. S., Phadia, E. G. and Tiwari, R. C. (1991) Bayesian nonparametric inference, *Current Issues in Statistical Inference; Essays in Honor of D.*

Basu (Ed. M. Ghosh and P. K. Pathak) IMS Lecture Notes - Monograph Series, **17** xxx—xxx.

Freedman, D. A. (1963) On the asymptotic behavior of Bayes estimates in the discrete case, *Ann. Math. Statist.* **34** 1194-1216.

Goldie, C. M. (1991) Implicit renewal theory and tails of solutions of random equations, *Ann. Applied Probab.* **1** 126-166

Harris, T. E. (1968) Counting measures, monotone random set functions, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **10** 102-119.

Kumar S., and Tiwari, R. C. (1989) Bayes estimation of reliability under a random environment governed by a Dirichlet prior, *IEEE Transactions on Reliability* **38** 218-223.

Sethuraman, J. and Tiwari, R. C. (1982) Convergence of Dirichlet measures and the interpretation of their parameter, *Statistical Decision Theory and Related Topics III* **2** 305-315.

Wilks, S. S. (1962) *Mathematical Statistics*, John Wiley & Sons, New York.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE				
4. PERFORMING ORGANIZATION REPORT NUMBER(S) FSU Department of Statistics Technical Report No. M-843			5. MONITORING ORGANIZATION REPORT NUMBER(S) D94603-90-6-0103	
6a. NAME OF PERFORMING ORGANIZATION Florida State University		6b. OFFICE SYMBOL (If applicable)		7a. NAME OF MONITORING ORGANIZATION U. S. Army Research Office
6c. ADDRESS (City, State, and ZIP Code) Department of Statistics B-167 Florida State University Tallahassee, FL 32306-3033			7b. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U. S. Army Research Office		8b. OFFICE SYMBOL (If applicable)		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER ARO 27468.7-MA
8c. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211			10. SOURCE OF FUNDING NUMBERS	
			PROGRAM ELEMENT NO.	PROJECT NO.
			TASK NO.	WORK UNIT ACCESSION NO.
11. TITLE (Include Security Classification) A Constructive Definition of Dirichlet Priors				
12. PERSONAL AUTHOR(S) Jayaram Sethuraman				
13a. TYPE OF REPORT Technical		13b. TIME COVERED FROM TO		14. DATE OF REPORT (Year, Month, Day) May, 1991
15. PAGE COUNT 12				
16. SUPPLEMENTARY NOTATION The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.				
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP	Bayesian Nonparametrics, Random Probability Measures, Dirichlet Measures.	
19. ABSTRACT (Continue on reverse if necessary and identify by block number) SEE BACK.				
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL			22b. TELEPHONE (Include Area Code)	
			22c. OFFICE SYMBOL	

The "parameter" in a Bayesian nonparametric problem is the unknown distribution P of the observation X . A Bayesian uses a prior distribution for P , and after observing X , solves the statistical inference problem by using the posterior distribution of P , which is the conditional distribution of P given X . For Bayesian nonparametrics to be successful one needs a large class of priors for which posterior distributions can be easily calculated.

Unless X takes values in a finite space, the unknown distribution P varies in an infinite dimensional space. Thus one has to talk about measures in a complicated space like the space of all probability measures on a large space. This has always required a more careful attention to the attendant measure theoretic problems.

A class of priors known as Dirichlet measures have been used for the distribution of a random variable X when it takes values in \mathcal{R}_k , see Freedman (1963), Fabius (1964) and Ferguson (1973). This family forms a conjugate family and possesses many pleasant properties.

In this paper we give a simple and new constructive definition of Dirichlet measures and remove the restriction that the basic space should be \mathcal{R}_k . We give complete self contained proofs of the three basic results for Dirichlet measures:

1. The Dirichlet measure is a probability measure of on the space of all probability measures,
2. it gives probability one to the subset of discrete probability measures, and
3. the posterior distribution is also a Dirichlet measure.